

Chapter 10

Next-Generation and Third-Generation Sequencing of Lung Cancer Biomarkers

Bryce Portier

Introduction

It is well established that lung cancer is an aggressive disease and it remains the leading cause of cancer-related deaths. Worldwide, there are more than 1.8 million new lung cancer cases diagnosed annually and over 1.5 million lung cancer-related deaths [1]. Morphologically, lung cancer is subdivided into two main types: non-small-cell lung cancer (NSCLC) which accounts for the majority 85% of new diagnosis and small-cell lung cancer which accounts for the minority 15% [2]. While NSCLC cases represent the majority of new diagnosis, this group can be further subdivided by morphologic and immunotypic methods into squamous cell carcinoma (SCC), adenocarcinoma (ADC), or large-cell lung carcinoma [3]. If all subtypes and clinical stages of lung cancer are combined, only 16% of patients achieve a benchmark of 5-year survival, which is largely due to the late stage (advanced disease progression) at the time of initial diagnosis [4]. In cases detected at an early stage (still localized), the 5-year survival rate is greatly increased to approximately 53% [5]. In an effort to improve early disease detection, new National Comprehensive Cancer Network (NCCN) guidelines recommend increasing low-dose computerized tomography (CT) screening. Furthermore, potential new diagnostic assays such as the automated three-dimensional morphologic analysis of epithelial cells in sputum (LuCED lung test) hopefully will increase the detection of early stage lung cancers [6, 7].

In addition to early detection, advances in understanding lung tumor biology and genomics are aiding in discovering new effective treatment solutions. Understanding the mechanisms and pathways that drive oncogenesis has directly led to the discovery of two predictive biomarkers in lung cancer: (1) epidermal growth factor receptor (*EGFR*) and (2) anaplastic lymphoma kinase (*ALK*) [8]. For *EGFR*, multiple

B. Portier, MD, PhD
Medical Affairs, Roche Tissue Diagnostics, Tucson, AZ, USA
e-mail: bryce.portier@roche.com

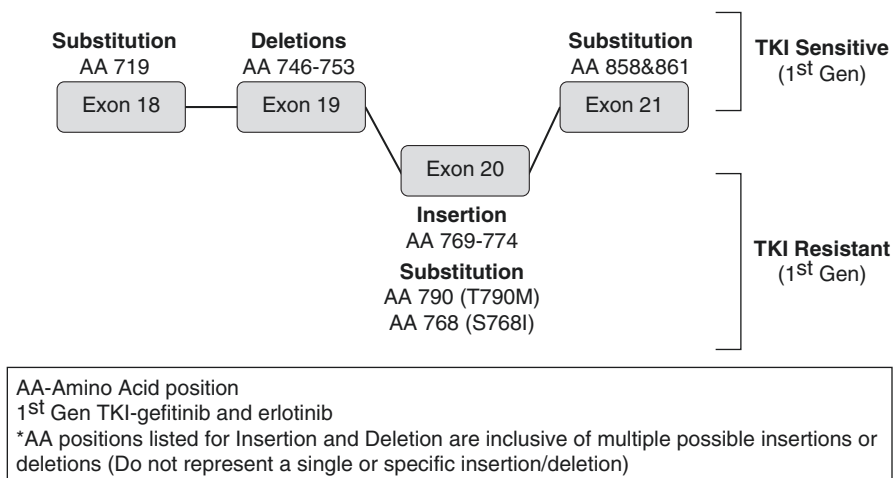


Fig. 10.1 Correlation of *EGFR* mutations and predicted TKI response

clinically significant alterations are known to occur in exons 18–21. Depending on the specific *EGFR* mutation, selection of a specific targeted therapy with sensitivity for that mutation can be determined (Fig. 10.1). One technology leading the path for new biomarker discovery and identification of driver pathways in lung cancer is next-generation sequencing (NGS). In the clinic, NGS technology is playing an essential role in interrogating large numbers of patients and screening vast portions of the genome in the search for altered genetic pathways and driver alterations in lung cancer. It is hopeful that NGS-based techniques, paired with prospective clinical trials, will expand our lung cancer biomarker knowledge and biomarker menu. Currently, only a limited set of biomarkers are routinely utilized in lung cancer clinical screening for targeted therapy selection (*EGFR*, *ALK*, and *ROS1*). In the following section, we will explore NGS with a focus on its clinical utility/benefits and variety of methodologies for addressing specific clinical questions and discuss barriers to widespread clinical adoption.

Next-Generation Sequencing (NGS) Background

The clinical use of NGS has significant benefits for diagnostic biomarker discovery and clinical screening capacity compared to traditional molecular assays such as single-gene Sanger-based sequencing (also referred to as first-generation sequencing) (Table 10.1). For clarification, the terminology “next”-generation sequencing or “NGS” refers to sequencing methodologies other than the traditional first-generation Sanger di-deoxy sequencing. NGS broadly encompasses both currently utilized methods referred to as “second-generation sequencing”

Table 10.1 Summary of NGS testing benefits and potential barriers to clinical adoption

<i>Benefits of NGS vs. Sanger sequencing</i>
Screen multiple genes/samples at one time (massive parallel)
Low input DNA/RNA required
Sensitivity to detect mutant allele at 2–10% (Sanger 15–25%)
Sensitive variant detection in samples with heterogeneity
Quantitative assay
Lower cost per sample/target
Ability to detect copy number alterations and gene fusions
<i>Barriers to NGS adoption</i>
Multiple platforms and rapidly evolving technology
Upfront cost of instrument and training
Lack of guidelines and unclear on process of LDT validation
Complex workflow with need for bioinformatic expertise
Dedicated hardware for analysis and long-term data storage

and new advancements in sequencing known as “third-generation sequencing” technologies. While debate over the exact categories for second- and third-generation sequencing Next-generation sequencing (NGS) exist, in general, second-generation sequencing represents methods that amplify DNA via emulsion PCR (e.g., Ion Torrent) or solid-phase amplification (e.g., Illumina). These methods are in contrast with third-generation sequencing which is performed utilizing non-amplified, single molecules (e.g., Pacific Biosciences and Oxford Nanopore). Regardless of classification as “second-” or “third-”-generation sequencing, both methods are encompassed in the term “next-generation sequencing” in which the term “next” refers to any non-Sanger-based sequencing methodology.

The utility of NGS (second and third generation) over that of Sanger is the ability to perform massive parallel sequencing. In essence, massive parallel sequencing involves interrogation of numerous samples and numerous alterations with speed and accuracy. Ultimately, this results in higher throughput which reduces cost per sample. NGS is also highly flexible with specific applications that can be tailored to the clinical question [9]. The clinical use of NGS has fundamentally improved our understanding of lung cancer biology and has led to revolutionizing clinical molecular diagnostic testing. As diagnostic lung tissue is often limited, NGS allows interrogation of numerous targets with limited sample input secondary to its ultralow sample input requirement. It is also capable of detecting mutations below 15% mutant allele frequency (compared to Sanger which requires 15–25% mutant allele frequency) [10].

Input material for NGS can be either DNA or RNA. Multiple sequencing methods exist and include whole-genome sequencing (WGS for DNA), whole-exome sequencing (WES for DNA), whole-transcriptome sequencing (RNA-Seq for RNA), and targeted sequencing (TS either DNA or RNA). Each method WGS, WES, RNA-Seq, or TS has specific strengths and weaknesses. In general, DNA-based methods identify small base pair alterations, insertion/deletions, as well as potential copy number changes. One significant difference between WGS, WES, and TS is depth of sequencing reads generated per target, which is higher for TS assays which focus on a selection of targets that typically represent a small fraction of the exome or genome. For instance, in lung cancer, TS-based NGS assays could focus on known genomic alterations in key biomarkers. RNA-based sequencing is utilized for detection of alternative gene-spliced transcripts, posttranscriptional modifications, gene fusion, mutations/single-nucleotide polymorphisms, small and long noncoding RNAs, or changes in gene expression. These methods will be explored and described in more detail below.

NGS Methodology

Whole-Genome Sequencing (WGS)

Currently, WGS represents one of the highest cost NGS methods and is not routinely utilized in routine clinical screening or monitoring of lung cancer. However, like most technologies, the cost of WGS is declining with improvements in NGS technologies [9]. WGS can detect a wide range of genomic alterations, including known disease-associated and novel variants, a feature that makes this technique well suited for research. Barriers to routine clinical lung cancer screening include the cost, the large volume of data produced, and necessary expertise/tools for data mining. Data analysis is a significant challenge for WGS, and streamlined process needs to be generated for this method to fulfill the gaps needed in personalized medicine [11]. Clinical strengths of WGS include the ability to determine breakpoints in balanced chromosome translocations and inversions and detecting genomic alterations outside of coding regions [12]. WGS allows full interrogation of promoters, enhancers, introns, noncoding RNAs (i.e., miRNAs), and unannotated regions [13, 14]. This full view of the genomic landscape is well suited for research applications or driver pathway discovery where a comprehensive profile of point mutations, complex rearrangements, indels, and copy number alterations is required [12]. For example, The Cancer Genome Atlas (TCGA) Research Network utilized WGS for lung adenocarcinomas and identified 25 significantly mutant genes, including both known mutations, *TP53* (50%), *KRAS* (27%), *EGFR* (17%), *STK11* (15%), *KEAP1* (12%), *ATM*, *NF1* (11%), *BRAF* (8%), and *SMAD4* (3%), and unknown (never previously reported) mutations, *SMARCA4*, *ARID1A*, *RBM10*, *SETD2*, *PICK3CA*, *CBL*, *FBXW7*, *PPP2R1A*, *RBI*, *CTNNB1*, *U2AF1*, *KIAA0427*, *PTEN*, *BRD3*, *FGFR3*, and *GOPC* [15]. The trade-off for such complete genomic landscape

analysis is low sequencing coverage. This one feature greatly limits the clinical application for routine lung cancer screening. WGS coverages vary depending on methodology but on average are below 100-fold, whereas targeted sequencing assays routinely achieve greater than 1000-fold coverage. Fold coverage is directly correlated with ability to identify tumors with low mutation burden, which is especially problematic in tumors that are not clearly separated from non-tumor stroma (dilutes mutant allele burden) [10].

RNA Sequencing (RNA-Seq)

RNA-Seq is a specialized form of NGS which can be utilized to interrogate the lung cancer transcriptome (represents up to ~4% of the human genome) [16]. Following the central dogma of molecular biology, DNA is transcribed to messenger RNA (mRNA), and mRNA is translated into protein. While the human genome contains approximately 25,000 genes, not all genes will be transcribed and translated into protein. Moreover, not every coded gene will be transcribed in proper order due to alternate splicing. Therefore, sequencing RNA (specifically mRNA) allows one to address questions including what genes are being expressed and at what level of expression. RNA-Seq can generate a comprehensive profile of the complete transcriptome or be utilized for a more focused targeted sequencing application. RNA-Seq as a method allows mapping the boundaries of exons and introns for identification of splice variants, identification of gene translocations, posttranscriptional modification, mutations, and noncoding of miRNAs [9, 12, 17]. It also offers a highly sensitive assay for quantification of the abundance of a transcript, even higher than comparative microarray technology [18]. While RNA-Seq offers several options not available by DNA-based NGS, it has its own inherent challenges which include library construction (inherently more difficult due to labile RNA molecule), data mining (high number of low abundant transcripts—potential false-positive calls), and obtaining complete transcript coverage [19].

Whole-Exome Sequencing (WES)

WES is utilized to specifically sequence the coding exons (~2.5% of the human genome) or the portion of genes that form the template for mRNA and successive protein production. This methodology specifically ignores noncoding regions such as promoters, enhancers, introns, and noncoding RNAs. Elimination of sequencing in these regions decreases the number of sequencing targets and thereby allows for improved fold coverage. WES focus solely on coding exons in annotated genes and therefore only allows variant detection in known coding genes. WES can be designed to also include sequencing of selected or limited regions of noncoding DNA regions which include exon-flanking regions and potentially select miRNAs. Similar to

WGS, the amount of sequencing data can be extensive for each sample and the number of total detected variants by WES can be high (20,000–30,000 range) depending on tumor sample and NGS methods/bioinformatics utilized. This large number of variants makes detecting actionable activating mutations a challenge. While more focused than WGS, application of WES to lung carcinoma is still currently best suited for research rather than routine clinical practice. Improvements in NGS such as decreased cost, faster analysis time, increased coverage, and improved accuracy could drive increased adoption of WES into routine clinical practice [10].

Targeted Sequencing (TS)

TS represents the most clinically utilized current NGS assay for lung cancer diagnostic testing. This method focuses specifically on interrogation of known genomic regions of interest. TS limits the sequencing to a small number of targeted regions, ultimately decreasing the amount of sequencing time and data generated, while also making the assay highly cost-effective by increasing the number of samples that can be analyzed simultaneously (multiplexed). Limiting TS to known cancer-relevant alterations makes this assay highly suited for clinical use which requires detecting known alterations such as point mutations and deletions in *EGFR* or even translocations in *ALK* or *ROS1*. However, being so highly targeted, this method may miss variants that are present but not located in regions interrogated by the assay. The adoption of TS via NGS into clinical practice for lung cancer has resulted in the availability of a highly sensitive method for detecting actionable alterations in lung cancer specimens [20–22]. A recent report showed NGS-based TS was able to identify *EGFR/KRAS/ALK* alterations in up to 58% of patients that were called wild type by standard testing, which translated into improved opportunities for therapeutic intervention [23]. Since most NSCLCs are detected once locally advanced and/or inoperable tumors, often only fine needle aspirate (FNA) cytology samples of mets are available for molecular testing. FNA tumor cell content may be very limited and therefore testing by traditional Sanger sequencing would not be possible. However, TS via NGS can utilize nanogram quantities of DNA, and FNA/cytology samples have been shown to be sufficient for TS NGS analysis [24–26].

NGS Translocation Detection

Currently, the list of routinely tested and actionable translocations specific for lung cancer is small and includes *ALK*, *RET*, and *ROS1*. Other kinase gene fusions have been detected by NGS from isolated lung adenocarcinoma DNA and RNA and include *MPRIP-NTRK1*, *AXL-MBIP*, *SCA11-PDGFR*, and *EZR-ERBB4* [27–29]. Regardless of molecular methodology utilized for detection, accurate identification of translocations can be challenging. Utilizing in situ hybridization (ISH) is the

current gold standard, but immunohistochemistry (IHC) is often performed as it offers a faster and less burdensome screening/detection methodology. However, IHC does not actually identify the translocation; rather, it identifies overexpression of a protein that occurs secondary to the translocation. Therefore, the IHC approach is applicable for *ALK* which lacks endogenous expression in the lung, but is not a viable option for identification of *RET* translocations due to endogenous *RET* expression [30] and potentially not useful for *ROS1* due to false-positive staining and poor correlation with FISH [31]. Unlike ISH and IHC options, NGS can be applied to identify both known and de novo translocations. In addition, NGS allows the simultaneous screening of actionable gene fusions in a single assay with high specificity and low input requirements (sample preservation). The inherent difficulty in identifying translocations via NGS is the high variability of translocation partners and breakpoints along with low incidence of translocations in lung cancer. While the canonical *EML4-ALK* fusion consists of *EML4* exons 1–13 fused to *ALK* exons 20–29, over 20 different *ALK* translocation partners have been identified [32]. NGS is gaining clinical utilization for translocation detection in lung carcinoma due to its comprehensive screening of multiple low incidence translocations, paired with high sensitivity for detection, rapid assay run time, and lower cost compared to single assay/single translocation testing options such as ISH [33]. Ultimately, the goal of utilizing NGS for translocation detection is to properly and rapidly stratify patients to the proper best personalized targeted therapy (sunitinib, sorafenib, or vandetanib) [28, 34].

NGS Utilizing Liquid Biopsy

The overarching trend in molecular diagnostics is to do more with less. NGS is perfectly suited for this task, as very little material is required for testing and the methodology is flexible to allow full mutation profiling or translocation screening. However, this is only applicable when tissue or cytology samples are available, which is not the case for routine follow-up or disease management. In these cases, often minimally invasive blood draws (liquid biopsies) are performed. Recently, much interest is focused on nucleic acid isolation from liquid biopsies via capturing rare circulating tumor cells (CTCs) or cell-free DNA (CF-DNA). A detailed discussion on the advantages and disadvantages of CTCs vs. CF-DNA is outside the scope of this article; however, a good summary was recently published [35]. Both CTC and CF-DNA have been successfully applied to capture starting material for clinical NGS testing. CTCs have already shown utility for NGS-based *EGFR* mutation testing, with one study showing an 84% match in CTC *EGFR* mutation profile compared to tissue biopsy and in addition multiple *EGFR* mutations were identified demonstrating the possibility of detecting tumor heterogeneity [36]. Likewise, CF-DNA has been successfully utilized for NGS-based lung cancer diagnostic testing for both general mutation screening and focused identification of acquired tyrosine kinase inhibitor (TKI) resistance *EGFR* mutations [37, 38]. The difficulty with

)CTC or CF-DNA applications is the very limited amount of DNA and the mixture of genomic and tumor nucleic acid. To overcome these challenges, NGS methodologies have been developed such as Tagged Amplicon Deep Sequencing (TAM-Seq), Safe Sequencing System (Safe-SeqS), and Cancer Personalized Profiling by deep sequencing (CAPP-seq) which have demonstrated up to 92% sensitivity and >99.99% specificity for *EGFR* mutation detection at the variant level [39–42]. These novel NGS methods improve the sensitivity of standard NGS by performing highly targeted hybrid capture, high-throughput deep sequencing, and utilizing bioinformatic tools to remove artifacts and discover rare mutations and potentially translocations [43].

Barriers to Adoption of Clinical NGS for Lung Cancer

While NGS has gained widespread use as a research tool, it has only been in the last few years that it has started to gain acceptance and utilization in the highly regulated clinical CAP/CLIA laboratory-based environment. Several barriers exist for widespread clinical adoption including cost, rapid technology change, lack of regulatory guidance, and complex bioinformatic data interpretation challenges (Table 10.1). These items will be discussed in detail below.

Cost of Clinical NGS Testing

Like most new technologies, NGS instrumentation and reagents can represent a high-cost burden for labs interested in undertaking the task of starting NGS testing. Instrument prices vary from sub-100,000 US dollar benchtop sequences to over 1,000,000 US dollars for high-throughput instrumentation. On top of instrument capital purchase cost, there is an annual service contract (price is highly variable). There are also costs for reagents, assay validations, personnel, and data analysis. NGS has a high upfront and operation cost relative to other molecular diagnostic equipment such as real-time PCR or Sanger-based assays. Cost can be greatly minimized per sample or test by the high degree of multiplexing that is capable, but lab volume and in-house expertise should be considered before initiating a NGS sequencing assay in the clinical setting. An additional variable that should be considered is the amount of testing reimbursement that will be generated by NGS testing. Current Procedural Terminology (CPT) codes are continually updated and in 2017 CPT codes for NGS-based testing exist [44]. However, the rate of successful reimbursement and the amount of reimbursement can be highly variable depending on geographic location and payer. This uncertainty in financial return is a direct barrier to widespread clinical adoption.

Guidelines

Although NGS is extensively used for research, its application in clinical practice has not been fully realized in part due to the lack of formalized validation and testing guidelines. NGS testing, while a promising method for lung cancer screening, is still a relatively new technology, and therefore, standards for validation in the CAP/CLIA lab are not well established. In addition, the regulation of laboratory-developed tests (LDT) in general has been a major unanswered question. The Food and Drug Administration (FDA) issued draft guidance in 2014 outlining new enforcement of testing regulation specifically targeting LDTs [45]. Based on this draft guidance, it was unclear what regulation NGS-based LDT testing would follow. However, recently, the FDA released a white paper that stated it would not issue a final guidance on the oversight of LDT [46]. This publication has largely cleared the way for NGS assay validations to move forward and fall under the regulatory guidance of CAP inspections and inclusion in proficiency testing, similar to other high complexity assays performed in the CAP/CLIA clinical laboratory setting. Moving forward, NGS still represents a unique validation challenge for molecular diagnostic labs, while well acquainted with running DNA-/RNA-based assays “wet lab,” it is the post-run analytical component of NGS that is difficult to validate, due in part to the novelty of NGS analytic tools and novel skills required for NGS bioinformatic data. In addition, NGS analysis requires a multistep “pipeline” method for processing data in which small deviations in assay design or post-sequencing analytic processing (filters) can impose any number of potential downstream errors. Despite these challenges, NGS is still being adopted in academic and private hospitals and has proven to be a profitable entity for commercial companies [47].

Bioinformatics

An in-depth exploration of bioinformatic approaches utilized in clinical NGS analysis is outside the scope of this chapter. However, it is worth mentioning that the large-scale data produced by NGS is a significant obstacle to adoption of clinical-based NGS assays [48]. To identify variants from NGS data, often multiple software packages need to be stitched together into a data analysis pipeline. These programs include sequence aligners, variant callers, and variant annotation. Each software component allows modification of multiple variables that can be user altered to allow highly customized workflows but at the price of decreased standardization and ability to perform quality assessments between labs. Most “pipelines” will consist of a sequence aligner (maps sequencing reads to a reference genome), variant caller (identifies variant sites), and variant annotation (links variant calls to database with annotated lists of clinical variants such as Catalog for Somatic Mutations in Cancer (COSMIC) [49]. The performances of different aligners have been

extensively studied and each has pros/cons, making adoption of a uniform analysis pipeline unlikely [50, 51]. One common tool that offers a good introduction to NGS data analysis is the Genome Analysis Toolkit (GATK; Broad Institute, Cambridge, MA, USA) [49]. This toolkit allows multiple standardized forms of NGS analysis and has well-documented instructions for users.

In addition to software standardization, an additional hurdle with clinical NGS is the sheer volume of data produced. This can cause analytic bottlenecks, even with recent advances and lowered costs of processing power. An additional potential problem is long-term data storage in a CAP-/CLIA-approved manner. Types of files or recommended length of storage for NGS has not been standardized at this time. Lastly, how labs are reimbursed for complex NGS analysis and even how labs integrate NGS data into electronic health records is highly variable, with no set standardization or national guidance [52–54].

Conclusion

There is clear evidence that NGS can accurately identify clinically significant biomarkers for lung cancer, such as *EGFR* mutations and *ALK* rearrangements, and that this technology can help guide personalized targeted therapies [55]. NGS performed on lung biopsies or cytology specimens can identify both established and emerging biomarkers depending on selected targets for sequencing and analysis [8, 56]. Likewise, NGS performed on CTCs or CF-DNA can be utilized to identify biomarkers for guided therapy or follow patients for monitoring development of tyrosine kinase inhibitor (TKI) resistance (such as *EGFR* T790M). Ultimately, the significance of applying clinical NGS to lung cancer screening is its ability to simultaneously interrogate numerous biomarkers and rapidly/accurately direct patients to an approved efficacious targeted treatment. Ongoing exploratory research utilizing NGS will undoubtedly translate into discovery and validation of novel predictive biomarkers, which will ultimately translate into clinical NGS practice and improve lung cancer diagnosis and treatment. Furthermore, it is expected that NGS technology will continue to advance at an accelerated rate and that the tangible outcome of this will be the improvement in our understanding of causal genetic mutations/alterations in lung cancer and continued improvement in lung cancer treatment and outcome.

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015;65(2):87–108.
2. Sequist LV, Lynch TJ. EGFR tyrosine kinase inhibitors in lung cancer: an evolving story. *Annu Rev Med.* 2008;59:429–42.
3. Travis WD, Brambilla E, Burke AP, Marx A, Nicholson AG. WHO classification of tumours of the lung, pleura, thymus and heart. Lyon: International Agency for Research on Cancer; 2015.

4. Cancer facts and figures 2016: American Cancer Society; [January 2017]. Available from: <http://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2016.html>.
5. Wu K, Huang RS, House L, Cho WC. Next-generation sequencing for lung cancer. *Future Oncol.* 2013;9(9):1323–36.
6. Meyer MG, Hayenga JW, Neumann T, Katdare R, Presley C, Steinhauer DE, et al. The cell-CT 3-dimensional cell imaging technology platform enables the detection of lung cancer using the noninvasive LuCED sputum test. *Cancer Cytopathol.* 2015;123(9):512–23.
7. Wilbur DC, Meyer MG, Presley C, Aye RW, Zarogoulidis P, Johnson DW, et al. Automated 3-dimensional morphologic analysis of sputum specimens for lung cancer detection: performance characteristics support use in lung cancer screening. *Cancer Cytopathol.* 2015;123(9):548–56.
8. Cagle PT, Raparia K, Portier BP. Emerging biomarkers in personalized therapy of lung cancer. *Adv Exp Med Biol.* 2016;890:25–36.
9. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet.* 2010;11(10):685–96.
10. Portier BP, Kanagal-Shamanna R, Luthra R, Singh R, Routbort MJ, Handal B, et al. Quantitative assessment of mutant allele burden in solid tumors by semiconductor-based next-generation sequencing. *Am J Clin Pathol.* 2014;141(4):559–72.
11. Diamandis M, White NM, Yousef GM. Personalized medicine: marking a new epoch in cancer patient management. *Mol Cancer Res.* 2010;8(9):1175–87.
12. Tuna M, Amos CI. Genomic sequencing in cancer. *Cancer Lett.* 2013;340(2):161–70.
13. Beck CR, Collier P, Macfarlane C, Malig M, Kidd JM, Eichler EE, et al. LINE-1 retrotransposition activity in human genomes. *Cell.* 2010;141(7):1159–70.
14. Huang CR, Schneider AM, Lu Y, Niranjana T, Shen P, Robinson MA, et al. Mobile interspersed repeats are major structural variants in the human genome. *Cell.* 2010;141(7):1171–82.
15. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature.* 2014;511(7511):543–50.
16. Pertea M. The human transcriptome: an unfinished story. *Genes (Basel).* 2012;3(3):344–60.
17. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* 2008;18(9):1509–17.
18. Morrissy AS, Morin RD, Delaney A, Zeng T, McDonald H, Jones S, et al. Next-generation tag sequencing for cancer gene expression profiling. *Genome Res.* 2009;19(10):1825–35.
19. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* 2009;10(1):57–63.
20. Masago K, Fujita S, Muraki M, Hata A, Okuda C, Otsuka K, et al. Next-generation sequencing of tyrosine kinase inhibitor-resistant non-small-cell lung cancers in patients harboring epidermal growth factor-activating mutations. *BMC Cancer.* 2015;15:908.
21. Rangachari D, VanderLaan PA, Le X, Folch E, Kent MS, Gangadharan SP, et al. Experience with targeted next generation sequencing for the care of lung cancer: insights into promises and limitations of genomic oncology in day-to-day practice. *Cancer Treat Commun.* 2015;4:174–81.
22. Shao D, Lin Y, Liu J, Wan L, Liu Z, Cheng S, et al. A targeted next-generation sequencing method for identifying clinically relevant mutation profiles in lung adenocarcinoma. *Sci Rep.* 2016;6:22338.
23. Lim SM, Kim EY, Kim HR, Ali SM, Greenbowe JR, Shim HS, et al. Genomic profiling of lung adenocarcinoma patients reveals therapeutic targets and confers clinical benefit when standard molecular testing is negative. *Oncotarget.* 2016;7(17):24172–8.
24. Scarpa A, Sikora K, Fassan M, Rachiglio AM, Cappellesso R, Antonello D, et al. Molecular typing of lung adenocarcinoma on cytological samples using a multigene next generation sequencing panel. *PLoS One.* 2013;8(11):e80478.
25. Qiu T, Guo H, Zhao H, Wang L, Zhang Z. Next-generation sequencing for molecular diagnosis of lung adenocarcinoma specimens obtained by fine needle aspiration cytology. *Sci Rep.* 2015;5:11317.

26. Treece AL, Montgomery ND, Patel NM, Civalier CJ, Dodd LG, Gulley ML, et al. FNA smears as a potential source of DNA for targeted next-generation sequencing of lung adenocarcinomas. *Cancer Cytopathol.* 2016;124(6):406–14.
27. Kohno T, Nakaoku T, Tsuta K, Tsuchihara K, Matsumoto S, Yoh K, et al. Beyond ALK-RET, ROS1 and other oncogene fusions in lung cancer. *Transl Lung Cancer Res.* 2015;4(2):156–64.
28. Lipson D, Capelletti M, Yelensky R, Otto G, Parker A, Jarosz M, et al. Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies. *Nat Med.* 2012;18(3):382–4.
29. Zheng Z, Liebers M, Zhelyazkova B, Cao Y, Panditi D, Lynch KD, et al. Anchored multiplex PCR for targeted next-generation sequencing. *Nat Med.* 2014;20(12):1479–84.
30. Lee SE, Lee B, Hong M, Song JY, Jung K, Lira ME, et al. Comprehensive analysis of RET and ROS1 rearrangement in lung adenocarcinoma. *Mod Pathol.* 2015;28(4):468–79.
31. Sholl LM, Sun H, Butaney M, Zhang C, Lee C, Janne PA, et al. ROS1 immunohistochemistry for detection of ROS1-rearranged lung adenocarcinomas. *Am J Surg Pathol.* 2013;37(9):1441–9.
32. Hallberg B, Palmer RH. Mechanistic insight into ALK receptor tyrosine kinase in human cancer biology. *Nat Rev Cancer.* 2013;13(10):685–700.
33. Pekar-Zlotin M, Hirsch FR, Soussan-Gutman L, Ilouze M, Dvir A, Boyle T, et al. Fluorescence in situ hybridization, immunohistochemistry, and next-generation sequencing for detection of EML4-ALK rearrangement in lung cancer. *Oncologist.* 2015;20(3):316–22.
34. Kohno T, Ichikawa H, Totoki Y, Yasuda K, Hiramoto M, Nammo T, et al. KIF5B-RET fusions in lung adenocarcinoma. *Nat Med.* 2012;18(3):375–7.
35. Calabuig-Farinas S, Jantus-Lewintre E, Herrerros-Pomares A, Camps C. Circulating tumor cells versus circulating tumor DNA in lung cancer-which one will win? *Transl Lung Cancer Res.* 2016;5(5):466–82.
36. Marchetti A, Del Grammastro M, Felicioni L, Malatesta S, Filice G, Centi I, et al. Assessment of EGFR mutations in circulating tumor cell preparations from NSCLC patients by next generation sequencing: toward a real-time liquid biopsy for treatment. *PLoS One.* 2014;9(8):e103883.
37. Couraud S, Vaca-Paniagua F, Villar S, Oliver J, Schuster T, Blanche H, et al. Noninvasive diagnosis of actionable mutations by deep sequencing of circulating free DNA in lung cancer from never-smokers: a proof-of-concept study from BioCAST/IFCT-1002. *Clin Cancer Res.* 2014;20(17):4613–24.
38. Murtaza M, Dawson SJ, Tsui DW, Gale D, Forshew T, Piskorz AM, et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature.* 2013;497(7447):108–12.
39. Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci U S A.* 2011;108(23):9530–5.
40. Lebofsky R, Decraene C, Bernard V, Kamal M, Blin A, Leroy Q, et al. Circulating tumor DNA as a non-invasive substitute to metastasis biopsy for tumor genotyping and personalized medicine in a prospective trial across all tumor types. *Mol Oncol.* 2015;9(4):783–90.
41. Martinez P, McGranahan N, Birkbak NJ, Gerlinger M, Swanton C. Computational optimisation of targeted DNA sequencing for cancer detection. *Sci Rep.* 2013;3:3309.
42. Newman AM, Lovejoy AF, Klass DM, Kurtz DM, Chabon JJ, Scherer F, et al. Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat Biotechnol.* 2016;34(5):547–55.
43. Newman AM, Bratman SV, To J, Wynne JF, Eclov NC, Modlin LA, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med.* 2014;20(5):548–54.
44. Icd-10-Cm the Complete Official Codebook: American Medical Association; 2017 ed. edition (September 30, 2016).
45. Administration FaD. Anticipated details of the draft guidance for industry, Food and Drug Administration Staff, and Clinical Laboratories. Framework for Regulatory Oversight of Laboratory Developed Tests (LDTs) 2014 [cited 2017 January 28th]. Available from: <http://www.fda.gov/downloads/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/UCM407409.pdf>.

46. Administration FaD. Discussion paper on laboratory developed tests (LDTs) 2017 [cited 2017 January 28th]. Available from: <http://www.fda.gov/downloads/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/LaboratoryDevelopedTests/UCM536965.pdf>.
47. Staff B-IW. Foundation Medicine, UnitedHealthcare sign lung cancer profiling agreement 2015 [cited 2017 January 28th]. Available from: <http://www.bio-itworld.com/brief/2015/12/22/foundation-medicine-unitedhealthcare-sign-lung-cancer-profiling-agreement.asp>.
48. Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. *Nat Rev Genet.* 2011;12(6):443–51.
49. Wellcome Trust Sanger Institute GRL. Catalogue of somatic mutations in cancer 2017. Available from: <http://cancer.sanger.ac.uk/cosmic>.
50. Costa JL, Sousa S, Justino A, Kay T, Fernandes S, Cirnes L, et al. Nonoptical massive parallel DNA sequencing of BRCA1 and BRCA2 genes in a diagnostic setting. *Hum Mutat.* 2013;34(4):629–35.
51. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature.* 2011;475(7356):348–52.
52. Hazin R, Brothers KB, Malin BA, Koenig BA, Sanderson SC, Rothstein MA, et al. Ethical, legal, and social implications of incorporating genomic information into electronic health records. *Genet Med.* 2013;15(10):810–6.
53. Kho AN, Rasmussen LV, Connolly JJ, Peissig PL, Starren J, Hakonarson H, et al. Practical challenges in integrating genomic data into the electronic health record. *Genet Med.* 2013;15(10):772–8.
54. Ury AG. Storing and interpreting genomic information in widely deployed electronic health record systems. *Genet Med.* 2013;15(10):779–85.
55. Campbell PJ, Stephens PJ, Pleasance ED, O’Meara S, Li H, Santarius T, et al. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat Genet.* 2008;40(6):722–9.
56. Cagle PT, Allen TC. Lung cancer genotype-based therapy and predictive biomarkers: present and future. *Arch Pathol Lab Med.* 2012;136(12):1482–91.